

# 基于改进 YOLOv5 的电厂人员吸烟检测

王彦生<sup>1</sup>, 曹雪虹<sup>1</sup>, 焦良葆<sup>1,2</sup>, 孙宏伟<sup>1</sup>, 高阳<sup>1</sup>

(1. 南京工程学院 人工智能产业技术研究院, 南京 211167;

2. 江苏省智能感知技术与装备工程研究中心, 南京 211167)

**摘要:** 发电厂厂区内违规吸烟易导致火灾、爆炸等事故, 会带来巨大损失; 针对电厂内人员违规吸烟行为检测精度不高的问题, 提出一种基于改进 YOLOv5s (You Only Look Once v5s) 的电厂内人员违规吸烟检测方法; 该方法以 YOLOv5s 网络为基础, 将 YOLOv5s 网络 C3 模块 Bottleneck 中的  $3 \times 3$  卷积替换为多头自注意力层以提高算法的学习能力; 接着在网络中添加 ECA (Efficient Channel Attention) 注意力模块, 让网络更加关注待检测目标; 同时将 YOLOv5s 网络的损失函数替换为 SIoU (Scylla Intersection over Union), 进一步提高算法的检测精度; 最后采用加权双向特征金字塔网络 (BiFPN, Bidirectional Feature Pyramid Network) 代替原先 YOLOv5s 的特征金字塔网络, 快速进行多尺度特征融合; 实验结果表明, 改进后算法吸烟行为的检测精度为 89.3%, 与改进前算法相比平均精度均值 (mAP, mean Average Precision) 提高了 2.2%, 检测效果显著提升, 具有较高应用价值。

**关键词:** 吸烟; 目标检测; 多头自注意力层; 注意力模块; 损失函数; 加权双向特征金字塔

## Smoking Detection of Power Plant Personnel Based on Improved YOLOv5

WANG Yansheng<sup>1</sup>, CAO Xuehong<sup>1</sup>, JIAO Liangbao<sup>1,2</sup>, SUN Hongwei<sup>1</sup>, GAO Yang<sup>1</sup>

(1. AI Industrial Technology Research Institute, Nanjing Institute of Technology, Nanjing 211167, China;

2. Jiangsu intelligent perception technology and equipment Engineering Research Center, Nanjing 211167, China)

**Abstract:** Illegal smoking easily leads to the accidents of fire and explosion in power plants, it will bring huge loss to the people. Aiming at the problem that the detection accuracy of illegal smoking behavior of personnel is not high in the power plants, an illegal smoking detection method of personnel is proposed on the basis of improved YOLOv5s in the power plants. The method is based on YOLOv5s network, the  $3 \times 3$  convolution in the C3 module of the YOLOv5s network with a Bottleneck layer is replaced to improve the learning ability of the algorithm. Then the efficient channel attention (ECA) module is added to the network, which makes the network pay more attention to the target to be detected. At the same time, the loss function of the YOLOv5s network is replaced by Scylla intersection over union (SIoU) to further improve the detection accuracy of the algorithm. Finally, the weighted bidirectional feature pyramid network (BiFPN) is used to replace original YOLOv5s feature pyramid network to rapidly perform multi-scale feature fusion. The experimental results show that, compared with the original algorithm, the detection accuracy of the improved algorithm is 89.3%, and the mean average precision (mAP) is increased by 2.2%. Its detection effect improves significantly, and it has high application value.

**Keywords:** smoking; target detection; multi-self attention layer; attention module; loss function; weighted bidirectional feature pyramid

## 0 引言

如今, 随着用电负荷的不断增长, 为了保证供电的可靠性, 发电厂的安全运行至关重要<sup>[1]</sup>。大多数发电厂内的设备仪器日常养护需要使用各种润滑油, 这些润滑油遇到明火极易引起火灾, 从而引发重大安全事故。因此需要规范约束工作人员的日常吸烟行为, 对违规吸烟行为进行监测警告。在日常生活中, 我们通常依靠烟雾报警器来监测吸烟行为, 该方法依靠传感器技术对空气进行实时采样,

尽管其检测精度较高, 但该方法无法准确判断是人为原因还是其他原因导致报警, 满足不了实时控烟的要求<sup>[2]</sup>。因此, 为了及时发现电厂中无人违规吸烟, 方便管理者对电厂内人员进行高效的管理, 有必要采用有效的目标检测方法, 通过视频监控电厂工作环境下的违规吸烟行为, 防止发生火灾火险。

近年来, 深度学习在目标检测领域的应用愈发广泛<sup>[3]</sup>。目前基于深度学习的目标检测方法主要分为两类, 一类是

收稿日期: 2022-09-21; 修回日期: 2022-10-25。

基金项目: 江苏省自然科学基金项目 (BK20201042); 江苏省政策引导类计划项目 (SZ2020007)。

作者简介: 王彦生 (1998-), 男, 甘肃临洮人, 硕士研究生, 主要从事深度学习、目标检测方向的研究。

曹雪虹 (1964-), 女, 江苏苏州人, 教授, 博士生导师, 主要从事无线通信系统、信息理论方向的研究。

焦良葆 (1972-), 男, 安徽安庆人, 教授, 硕士研究生导师, 主要从事图像信号处理、视觉信息理解方向的研究。

通讯作者: 王彦生 (1998-), 男, 甘肃临洮人, 硕士研究生, 主要从事深度学习、目标检测方向的研究。

引用格式: 王彦生, 曹雪虹, 焦良葆, 等. 基于改进 YOLOv5 的电厂人员吸烟检测[J]. 计算机测量与控制, 2023, 31(5): 48-55.

以 R-CNN (region-based convolutional neural network)、Fast R-CNN (Fast Region-based Convolutional Network)、Faster R-CNN (Faster Region-based Convolutional Network) 等为典型代表的双阶段目标检测算法<sup>[4]</sup>, 这种检测方法的主要思路是先预设一个可能包含待检测物体的预选框, 再通过卷积神经网络进行样本分类计算; 另一类则是以 YOLO (You Only Look Once) 系列、SSD<sup>[5]</sup> (Single shot multi-box detector) 等为典型代表的单阶段目标检测算法<sup>[6]</sup>, 与双阶段目标检测算法不同的是, 这种方法不用生成预选框, 而是直接在网络中提取特征值来分类目标和定位, 因此这种算法检测速度快, 检测精度也很高<sup>[7]</sup>。目前对于吸烟行为检测的目标检测方法主要以 Faster R-CNN 和 YOLO 系列为基础, 文献 [8] 通过检测人脸, 在人脸检测的基础上再利用 Faster R-CNN 算法检测烟支, 从而提高检测速度并降低误检率。文献 [9] 通过改进 YOLOv3 中的损失函数实现检测速度和精度的提升, 但该算法模型复杂度较高, 缺乏应用价值。文献 [10] 通过 YOLOv4 检测算法和人体关键点结合进行吸烟检测以提高检测精度, 尽管其在近距离条件下效果较好, 但是当人距离监控较远时误测严重。

针对吸烟行为检测精度较低的问题, 本文以 YOLOv5 6.0 版本中的 YOLOv5s 算法为基础, 通过修改网络结构和添加相关模块, 在保持网络轻量化的同时, 有效提升吸烟行为检测精度, 具有较高的实用价值。

### 1 YOLOv5 网络介绍

整个 YOLOv5 网络可以分为 4 个部分, 分别是输入端、骨干网络 Backbone、强特征提取网络 Neck 以及预测网络 Prediction<sup>[11]</sup>。YOLOv5 6.0 版本的具体结构如图 1 所示。

YOLOv5 6.0 版本的输入端与之前的 5.0 版本一样, 采用 Mosaic 数据增强、自适应锚框计算和自适应图片缩放。

6.0 版本在主干网络 Backbone 中与 5.0 版本有所不同, 6.0 版本使用一个卷积 (Convolution, Conv) 替换了 5.0 版本 Backbone 中的 Focus 层, 其目的是为了更方便模型导出。

此外还使用了 CSP 结构, 简称 C3 模块, 在 YOLOv5 中有两种 CSP 结构, CSP1\_X 结构应用于 Backbone 主干网络, 另一种 CSP2\_X 结构则应用于 Neck 中, CSP1\_X 将输入分为两个分支, 一个分支先通过 CBS (Conv+BN (Batch Normalization) + SiLU (Sigmoid Weighted Liner Unit)), 再经过多个残差结构 (Bottleneck×N), 另一个分支则只进行 CBS, 然后两个分支进行 Concat, 最后再进行一个 CBS。CSP2\_X 与 CSP1\_X 唯一的区别是 CSP2\_X 用 CBS 替换了 CSP1\_X 中的残差结构。对比普通的 CBS, CSP 结构分为两个支路, 有支路就意味着特征的融合, 而 Concat 就可以更好的把不同支路的特征信息保留下来, 因此 CSP 的设计可以提取到更为丰富的特征信息。在 Backbone 的最后采用 SPPF (Spatial Pyramid Pooling Fast) 的结构, 即空间金字塔池化, SPPF 结构将输入串行通过多个 MaxPool 层, 从而融合不同感受野的特征图, 丰富特征图的表达能力, 进一步提高运行速度。

在网络的 Neck 部分, 沿用了 FPN (Feature Pyramid Network) + PAN (Path Aggregation Network) 的结构, FPN 结构以自顶向下的方式构建出高级语义特征图, PAN 又加入了自底向上的路线, 弥补并加强了定位信息<sup>[12]</sup>。此外, 使用 CBS 结构来进一步加强网络特征融合。

预测网络 Prediction 的工作就是对特征点进行判断, 判断特征点是否有物体与其对应。

### 2 改进策略

#### 2.1 多头注意力机制

UC Berkeley 和 Google 基于 Transformers 结构设计了一种结构简单、功能强大的 Backbone, 名为 BoTNet<sup>[13]</sup>, 其改变是将 ResNet (Residual Network) 中的 3×3 卷积替换为多头注意力层 (Multi-Head Self-Attention, MHSA), 其具体结构如图 2 (a) (b) 所示, 具体的 MHSA 层如图 3 所示。

由上图 3 可知, MHSA 的输入尺寸为  $H \times W \times d$ , 分别表示输入特征矩阵的高宽以及单个 token 的维度, token 的数量即为  $H \times W$ 。首先是初始化两个可学习的参数向量  $R_h$ 、

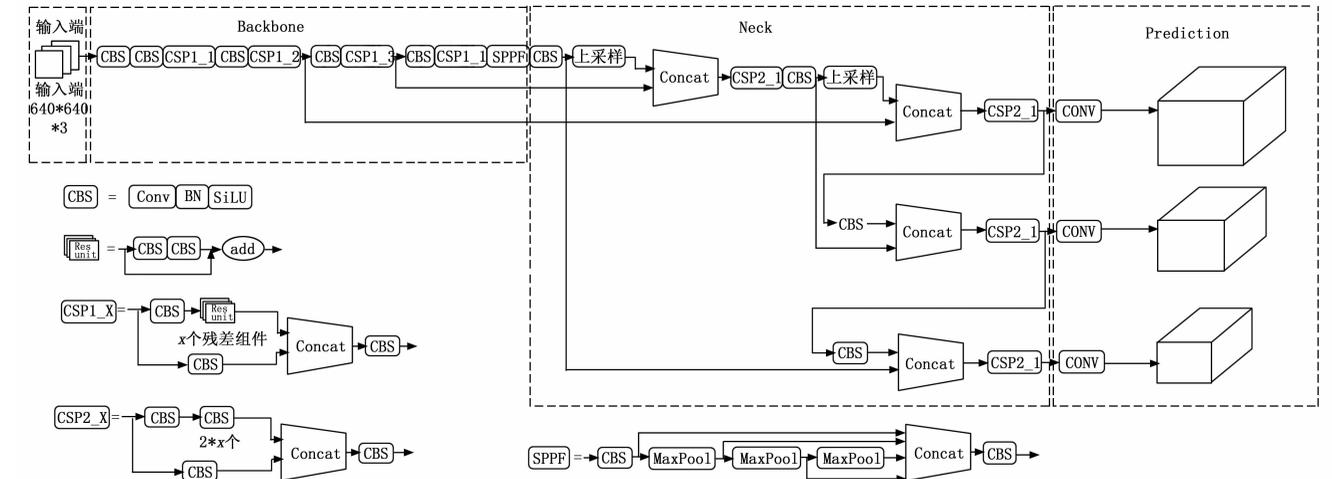


图 1 YOLOv5 6.0 版本的网络结构图

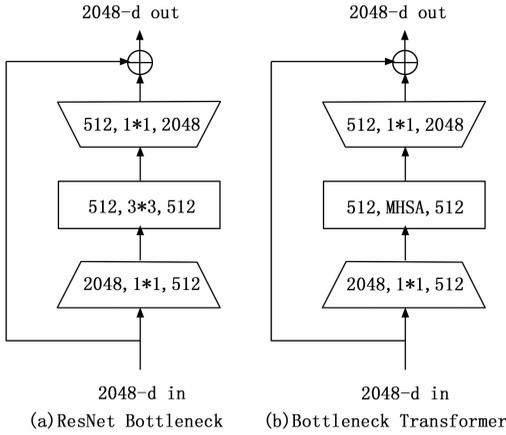


图 2 ResNet Bottleneck 与 BoTNet 网络结构

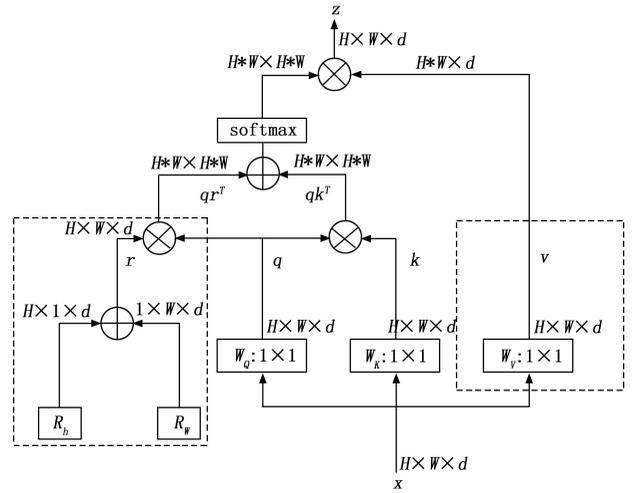


图 3 MHSA 层网络结构

$R_w$ ，分别表示高度和宽度不同位置的位置编码，然后将它们通过广播机制加起来，即位置的编码为  $R_{hi} + R_{wj}$  的两个  $d$  维向量相加，这样做将  $H \times W \times d$  个编码简化为  $(H+W) \times d$ 。此外，位置编码也不是直接加到输入上，而是与 query 矩阵进行矩阵乘法得到 Attention 的一部分，从而可以从多个维度提炼特征信息，最后将其与 query 和 key 算出来的加和后经过 softmax 得到最终的 Attention。

卷积神经网络 YOLOv5 通过堆叠卷积层来放大感受野，进而增强网络性能，但是卷积神经网络只关注图像局部，

无法兼顾图像全局。为此本文将 YOLOv5s 网络 Backbone 和 Neck 中 C3 模块的  $3 \times 3$  卷积替换为 MHSA，使 YOLOv5s 网络可以关注图像全局，并且能够基于相同的注意力机制学习到不同的行为，提高算法的学习能力，进而提高检测精度<sup>[14]</sup>。由于 Backbone 和 Neck 中 C3 模块 Bottleneck 结构不一样，因此图 4 分别展示了两个部分具体的替换过程，将替换后的结构命名为 BottleneckTransformer，最终替换后的结构如图 4 (d) 和 (h) 所示，替换后的结构

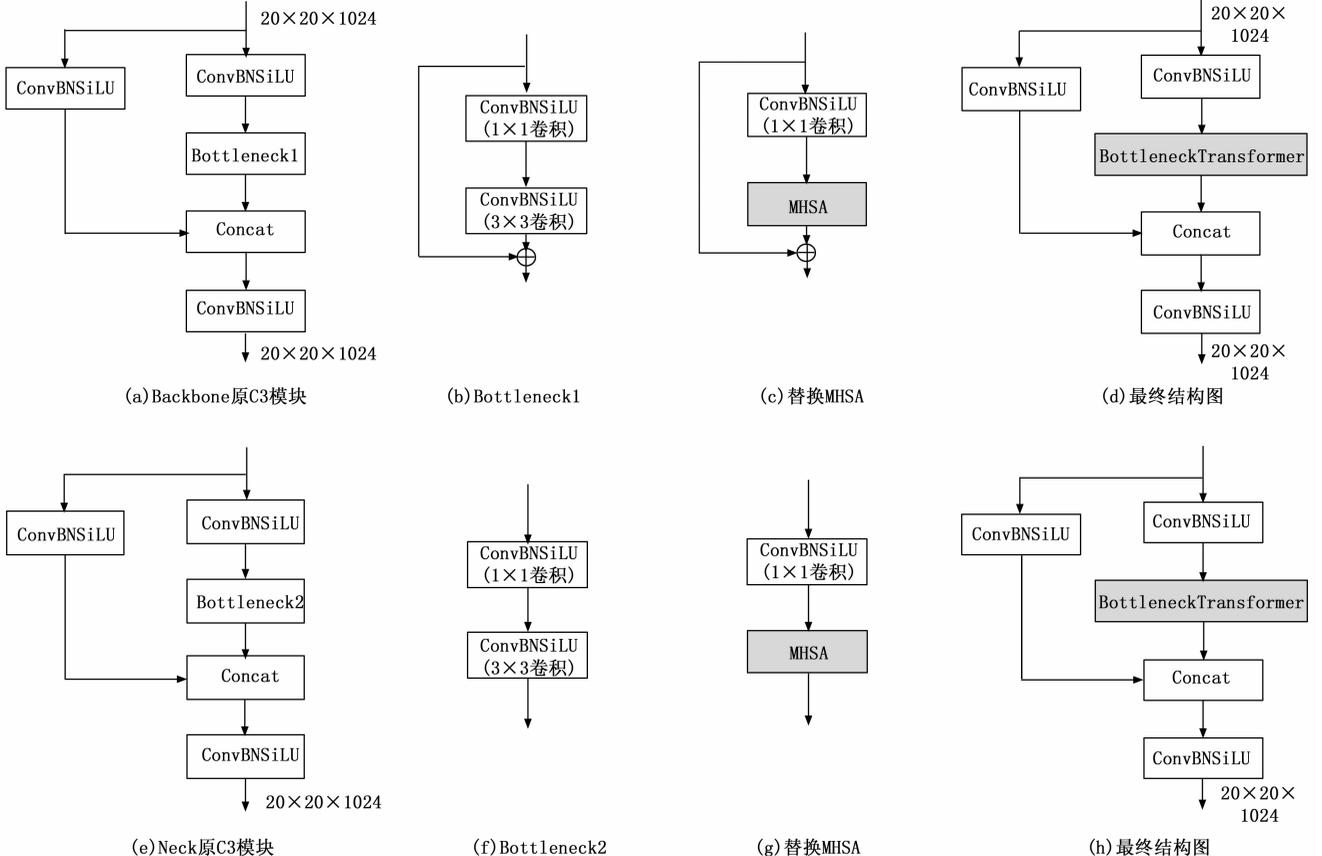


图 4 MHSA 的替换过程

与原网络 C3 模块相比, 唯一不同的就是将 Bottleneck 模块替换成了 BottleneckTransformer。

## 2.2 添加注意力机制 ECA

传统的注意力机制有很多, 例如 SE (Squeeze-and-Excitation)、CA (CoordAttention) 以及 CBAM (Convolutional Block Attention Module), 加入注意力机制的目的就是为了从众多信息中选择出当前任务更关注的信息。2020 年天津大学 Wang Qilong<sup>[15]</sup> 等人提出了一种高效的通道注意力机制 (ECA), 具体结构如图 5 所示, 作者通过对注意力机制 SE 的研究发现, SE 注意力机制首先会对输入特征图进行了通道压缩, 而这样的压缩降维对于学习通道之间的依赖关系有不利影响, 基于这个缺点, 提出了一种无需降维的局部跨通道交互策略, 该策略可以通过 1D 卷积有效地实现, 并且可以自适应选择一维卷积核的大小<sup>[16]</sup>。

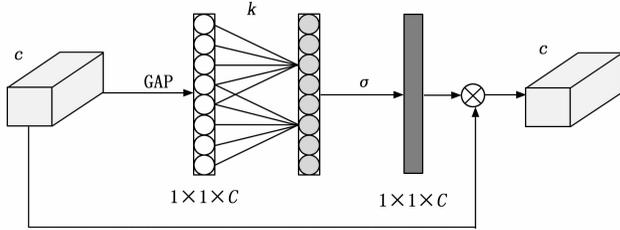


图 5 ECA 结构图

ECA 模块的实现主要分为三步, 第一步是将输入特征图进行全局平均池化 (Global average pooling, GAP) 操作, 从而获得未降维的所有特征; 第二步则进行卷积核大小为  $k$  的 1 维卷积操作, 以此来捕获局部跨通道交互信息, 接着经过 Sigmoid 激活函数得到各个通道的权重, 如公式 (1) 所示; 最后将权重与原始输入特征图对应元素相乘, 得到最终输出特征图<sup>[17]</sup>。

$$\omega = \sigma(C1D_k(y)) \quad (1)$$

式中, C1D 代表一维卷积;  $k$  为卷积核大小;  $y$  代表输入特征图;  $\sigma$  代表 Sigmoid 激活函数。

有时由于距离远, 从视频监控中监测到的吸烟人员相对较小, 使用原始的 YOLOv5s 算法无法检测到吸烟行为。为此本文引入 ECA 注意力模块, 通过将此模块添加在 YOLOv5s 网络 Backbone 部分中  $80 \times 80 \times 256$  的 CSPLayer 层输出位置, 以此来增强算法的特征学习能力, 让算法关注更有用的信息, 最终提高 YOLOv5s 算法对吸烟行为的检测效果。

## 2.3 修改损失函数

传统的目标检测损失函数 (例如 GIoU (Generalized Intersection over Union)、CIoU (Complete Intersection over Union)、IIoU (Improved CIoU) 等) 依赖于边界框回归度量的聚合, 如预测框和地面真值框的距离、重叠面积和长宽比。但是, 这种方法有一个明显的缺点, 那就是没有考虑真实框与预测检测框之间方向不匹配的问题, 使得目标检测算法在模型训练的过程中出现预测框“四处游荡”

的情况, 导致收敛速度较慢且效率较低, 最终产生更差的模型<sup>[18]</sup>。针对上述缺点, Zhora Gevorgyan<sup>[19]</sup> 提出了一种新的损失函数 SIoU, SIoU 充分考虑了所需回归之间的向量角度, 并且重新定义了惩罚指标, 从而有效提高模型训练的速度和推理的准确性。

SIoU 损失函数由 4 个成本函数组成, 分别是角度成本、距离成本、形状成本以及 IoU (Intersection over Union) 成本。在目标检测一开始的训练中, 大多数的预测框与真实框是不相交的, 为了快速的收敛两框之间的距离, 添加角度感知 LF 组件, 定义如公式 (2) 所示, 进而可以最大限度地减少与距离相关的变量数量。

$$\Delta = 1 - 2 \times \sin^2 \left( \arcsin(\sin(\alpha)) - \frac{\pi}{4} \right) \quad (2)$$

式中,  $\alpha$  为预测框与真实框之间的水平夹角。

但是直接使用角度损失是不合理, 因此还需要考虑距离。定义距离成本时考虑了公式 (2) 定义的角度成本。

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma_t}) \quad (3)$$

式中,  $\gamma$  为 2 减去角度成本, 当  $t$  为  $x$  时,  $\rho_x$  表示真实框和预测框宽的差值比上真实框和预测框最小外接矩形宽的平方值; 当  $t$  为  $y$  时,  $\rho_y$  表示真实框和预测框高的差值比上真实框和预测框最小外接矩形高的平方值。

形状成本定义为:

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega_t})^\theta \quad (4)$$

式中, 当  $t$  为  $w$  时,  $\omega_w$  表示预测框和真实框的宽的插值的绝对值比上两宽中的最大值; 当  $t$  为  $h$  时,  $\omega_h$  表示预测框和真实框的高的插值的绝对值比上两高中的最大值。 $\theta$  值定义形状成本, 其值对于每个数据集都是唯一的, 若  $\theta=1$ , 则代表立刻优化形状, 从而影响形状的自由运动。

IoU 表示预测框与真实框相交的面积比上相并的面积。最后 SIoU 损失函数定义如公式 (5) 所示:

$$L_{\text{box}} = 1 - \text{IoU} + \frac{\Delta + \Omega}{2} \quad (5)$$

原 YOLOv5s 网络中的损失函数为 CIoU 将边界框的纵横比作为惩罚项加入到边界框损失函数中, 一定程度上可以加快预测框的回归收敛过程, 但是一旦收敛到预测框和真实框的宽和高呈现线性比例时, 就会导致预测框回归时的宽和高不能同时增大或者减少。而用 SIoU 替换 CIoU 就可以轻而易举的解决这个问题, 从而有效提高模型训练的速度和推理的准确性。

## 2.4 特征金字塔改进

原本的 YOLOv5s 网络采用 FPN + PAN 的金字塔结构, FPN 自顶向下将高层的强语义特征传递下来, 对整个金字塔进行增强, 不过只增强了语义信息, 对定位信息没有传递, 而 PAN 就是针对这一点, 在 FPN 的后面添加一个自底向上的金字塔, 对 FPN 补充, 将低层的强定位特征传递上去, 结构如图 6 (a) (b) 所示。

尽管这种结构有效加强了网络的特征融合能力, 但是却存在一个问题, 那就是 PAN 没有接收到原始的信息, 只

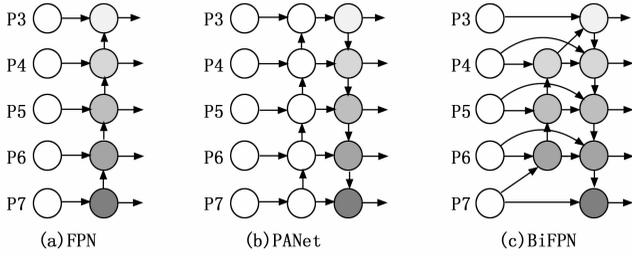


图 6 不同特征融合结构示意图

接收到 FPN 的输出信息，这就导致模型训练时目标检测算法学习出现偏差，进而导致检测精度低的问题，此外，带来很大的计算量。针对这个问题，谷歌提出一种新的双向特征网络 BiFPN，BiFPN 允许特征既能够自下而上也能够自上而下地反复流动，从而可以重复多次以得到更多高层特征融合，具体结构如图 7 (c) 所示<sup>[20]</sup>，与 FPN+PAN 结构相比，BiFPN 去掉了只有一条输入边和输出边的结点，并且如果输入和输出结点是同一层的，就额外添加一条边，如图 7 (c) 中的 P4、P5 和 P6 层。因此，用 BiFPN 网络替换 FPN+PAN 的金字塔结构，可以有效增强不同网络层之间特征信息的传递，明显提升 YOLOv5s 算法检测精度，并且具有不错的检测性能。

### 3 实验与分析

#### 3.1 数据集采集及制作

本文实验的数据集采用两种方式进行采集，第一种采集方式是先录制模拟电厂厂区内不同人员吸烟的视频，然后按照一定的帧率间隔截取视频获得图片，另一种采集方式是从网络上搜集不同场景下的吸烟图片，通过上述采集方法共采集图片 1 395 张，其中，视频截取图片 873 张，网络搜集图片 522 张。对于准备好的数据集，采用 LabelImg 软件进行标注。操作界面如图 7 所示，按照标签为 cigarette 进行标注，标注完的信息以 xml 的格式保存在相同路径下，然后通过代码脚本将 xml 文件转换为 YOLOv5 模型需要的 txt 文件。最后将数据集按 6 : 2 : 2 的比例划分为三部分，分别为训练集、验证集和测试集。

#### 3.2 实验环境与模型训练

本次实验的实验平台操作系统为 Ubuntu18.04，实验以 Pytorch 为软件框架，编程环境为 Python，CPU 为 Intel (R) Xeon (R) Glod5118CPU @ 2.3 GHz，GPU 为 GeForce RTX2080Ti，显存 24 G，内存 128 G。

实验参数的设置根据自己电脑的性能而定，由于本文实验使用有四张独显的实验室服务器，其性能远优于普通电脑，故其参数的设置相对较高。本文实验均是在不加载预训练权重的情况下进行，其他相关参数设置如表 1 所示。



图 7 LabelImg 软件操作界面

表 1 相关参数设置

参数	参数设置
输入图片大小	640×640
训练轮数	300
批次大小	48
交并比阈值	0.5
初始学习率	0.01

划分完数据集后，搭建 YOLOv5 网络环境，修改数据配置文件和模型配置文件，然后根据自己的电脑配置设置好相关的参数进行训练。训练结果如图 8 所示，从图中可以看出，box\_loss 和 obj\_loss 曲线逐渐下降，最后趋于稳定，这两条曲线趋近的值越小，表明方框越准，目标检测越准；cls\_loss 曲线指推测为分类 loss 均值，越小分类越准；Precision 曲线、Recall 曲线以及 mAP 曲线随着训练轮数的增加，逐渐趋于 0.9，表明训练效果较好，训练精度较高。

训练结束后会生成相应的权重文件，将测试代码中的权重路径修改为生成的 best 权重文件，其余相关参数设置同训练设置一样，然后运行代码就会生成测试结果。改进前后的网络对吸烟行为的检测如图 9 所示，图 (a) 表示原 YOLOv5s 的检测效果，图 (b) 表示改进后 YOLOv5s 算法

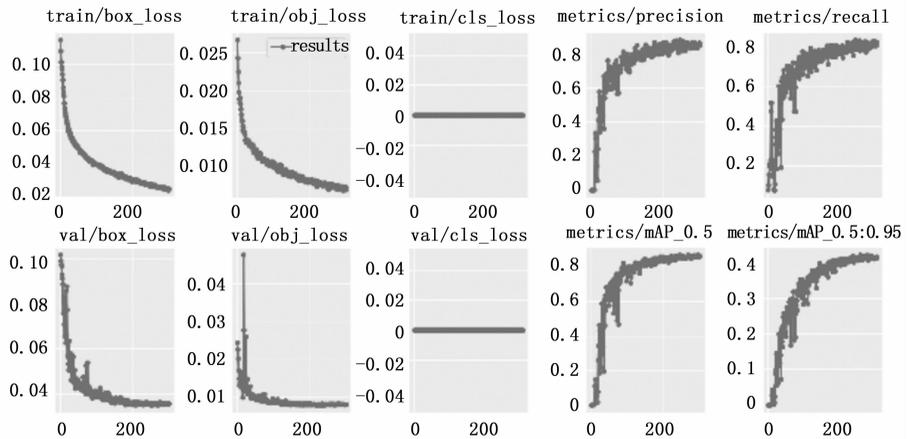


图 8 训练参数变化图

的检测效果, 可以看出, 使用原 YOLOv5s 检测时存在一定的误报和漏报, 而使用改进后的算法进行检测时, 误报和漏报情况明显改善, 并且对于视频监控下远距离的人员吸烟行为检测置信度达 79%, 属于较易检测出的对象, 检测效果较好。

换为多头自注意力层时, 检测精度提高 2.3%, 检测效果最好。

表 2 具体替换方案

方案	替换位置
A	位置 2、位置 4
B	位置 1、位置 3
C	位置 2
D	位置 1



(a) 原 YOLOv5s 检测效果 (b) 改进 YOLOv5s 检测效果

图 9 算法改进前后的检测效果对比图

表 3 替换多头自注意力层验证实验

算法	模型参数量/万	权重文件大小/MB	查准率/%	查全率/%	mAP@0.5/%	FPS
原 YOLOv5s	701	14.4	82.5	84.7	87.1	256
方案 A	635	13.1	87.6	85.0	89.4	181
方案 B	694	14.3	86.5	81.6	87.2	125
方案 C	668	13.7	87.2	83.5	87.8	185
方案 D	703	14.5	85.8	84.3	88.0	129

### 3.4.2 注意力模块实验

为了验证加入注意力模块后算法的检测能力以及探究应该将注意力模块加到算法的具体哪个位置, 本文通过在同一位置分别加入 ECA、CBAM、CA 三种注意力机制实验, 实验结果如表 4 所示, 实验结果表明, 加入 ECA 注意力机制的效果最好。然后再变换 ECA 加入的位置, 具体替换位置如图 10 所示, 具体实验结果如表 5 所示。通过实验得出, 将 ECA 加到 Backbone 部分中  $80 \times 80 \times 256$  的 CS-PLayer 层输出位置的效果最好, 精度提升 3.9%。

### 3.3 模型评价指标

目标检测实验评价中将模型参数量 (万)、模型权重文件大小 (MB)、查全率 (Recall, R)、查准率 (Precision, P)、平均精度均值 (mAP)、每秒检测帧数 (Frames Per Second, FPS) 这几项性能指标评判算法的性能。其中 P、R 和 mAP 分别为<sup>[21]</sup>:

表 4 融合不同注意力模块对比实验

算法	模型参数量/万	权重文件大小/MB	查准率/%	查全率/%	mAP@0.5/%	FPS
原 YOLOv5s	701	14.4	82.5	84.7	87.1	256
YOLOv5s+ECA	701	14.4	89.1	83.4	91.0	232
YOLOv5s+CBAM	720	14.8	83.3	83.7	87.1	133
YOLOv5s+CA	718	14.8	88.6	80.6	86.4	108

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$R = \frac{TP}{TP + FN} \quad (7)$$

$$mAP = \frac{\sum \frac{\sum P_n}{\sum r}}{num\_classes} \quad (8)$$

其中: TP (True Positives) 表示被正确检测出的目标数, FP (False Positives) 表示检测错误的目标数, FN (False Negatives) 表示未被检测出来的目标数,  $\sum r = 1$ ,  $\sum P_n$  表示 PR 曲线上时对应 P 的数值, num\_classes 为类别数。

表 5 融合注意力模块验证实验

算法	模型参数量/万	权重文件大小/MB	查准率/%	查全率/%	mAP@0.5/%	FPS
原 YOLOv5s	701	14.4	82.5	84.7	87.1	256
位置 5	701	14.4	85.1	82.4	87.9	136
位置 6	701	14.4	90.3	79.3	87.5	140
位置 7	701	14.4	89.1	83.4	91.0	232
位置 8	701	14.4	86.5	84.2	88.4	256

### 3.4 实验评估与分析

#### 3.4.1 替换多头注意力机制的实验分析

为了确定替换多头注意力机制的位置, 本文通过将不同位置的 C3 模块 Bottleneck 中的  $3 \times 3$  卷积替换成多头自注意力层, 在不加载预训练权重的情况下在 A、B、C、D 四种替换方案下进行实验, 替换位置如图 10 所示, 替换方案如表 2 所示, 实验结果如表 3 所示。通过实验最终确定将 Backbone 和 Neck 中两个 C3 模块 Bottleneck 的  $3 \times 3$  卷积替

#### 3.4.3 损失函数实验

为了验证 SiIoU 损失函数的有效性, 本文以 YOLOv5s 为基础, 通过替换原始网络损失函数为本文引入的 SiIoU 以及 EIou 和 Alpha-IoU 两种常用损失函数, 其他部分均不变, 在同一数据集上进行实验, 实验结果如表 6 所示。通

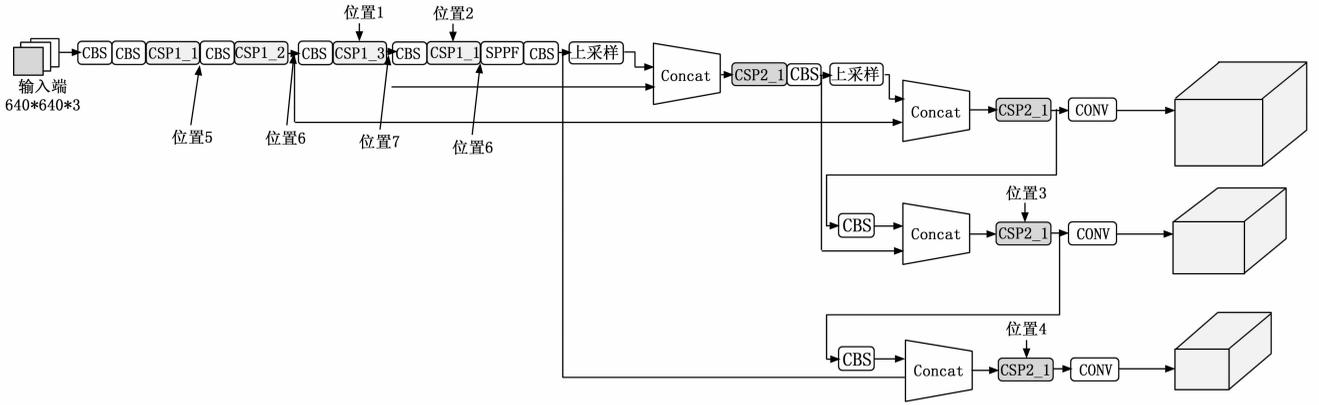


图 10 替换位置

过实验可以看出，相较于其他两个损失函数，原始网络替换完 Siou 的效果最好，精度提升了 3.6%。

表 6 改进损失函数 Siou 验证实验

算法	模型参数量/万	权重文件大小/MB	查准率/%	查全率/%	mAP@0.5/%	FPS
原 YOLOv5s	701	14.4	82.5	84.7	87.1	256
YOLOv5s+Siou	701	14.4	90.1	82.3	90.7	232
YOLOv5s+EIoU	701	14.4	84.0	83.1	87.1	181
YOLOv5s+Alpha-IoU	701	14.4	80.0	79.5	82.9	384

### 3.4.4 金字塔实验

为了验证改进 BiFPN 的有效性，本文将原始 YOLOv5s 算法和将 FPN+PAN 结构改为本文提出的 BiFPN 结构进行实验，实验结果如表 7 所示<sup>[22]</sup>。通过实验结果可以看出，替换完 BiFPN 以后，目标检测算法的检测精度提高了 0.9%。

### 3.4.5 消融实验

本文提出的 4 种改进方法分别为 MHSA、ECA、Siou、BiFPN。为了验证这 4 种不同改进方法的有效性，本文在原始 YOLOv5s 网络的基础上，分别只改进一种方法，然后再将 4 个改进点一同加入进行实验，实验结果如表 8 所示。实验结果表明，每种改进方法都显著提高了模型的检测精度，其中加入 ECA 注意力机制的精度提升最高，提高了 3.9%。此外，将 4 个改进点一同加入，精度提升也十分明显，相交原始 YOLOv5s 网络提高 2.2%。通过消融实验，再次证明了本文提出的改进点有效提高了目标检测算法的检测精度。

表 7 替换 BiFPN 验证实验

算法	模型参数量/万	权重文件大小/MB	查准率/%	查全率/%	mAP@0.5/%	FPS
原 YOLOv5s	701	14.4	82.5	84.7	87.1	256
YOLOv5s+BiFPN	808	16.6	84.8	81.8	88.0	204

表 8 消融实验结果

算法	模型参数量/万	权重文件大小/MB	查准率/%	查全率/%	mAP@0.5/%	FPS
YOLOv5s	701	14.4	82.5	84.7	87.1	256
YOLOv5s+MHSA	635	13.4	87.6	85.0	89.4	181
YOLOv5s+ECA	701	14.4	89.1	83.4	91.0	232
YOLOv5s+Siou	701	14.4	90.1	82.3	90.7	232
YOLOv5s+BiFPN	807	16.6	84.8	81.8	88.0	204
YOLOv5s+MHSA+ECA	653	13.4	88.4	80.2	87.1	133
YOLOv5s+MHSA+ECA+Siou	653	13.4	83.2	80.8	86.4	175
YOLOv5s+MHSA+ECA+Siou+BiFPN	742	15.3	87.7	83.3	89.3	227

### 3.4.6 对比实验

为了进一步证明本文改进算法的优越性和有效性，有必要将其与其他目标检测算法进行对比，因此将本文算法与传统的 YOLOv3、YOLOv3-tiny、YOLOv5l、YOLOv5m、YOLOv5x、YOLOv5s 以及 YOLOv5n 进行对比，实验结果如表 9 所示。

表 9 不同的算法模型性能对比

算法	模型参数量/万	权重文件大小/MB	查准率/%	查全率/%	mAP@0.5/%	FPS
YOLOv3	6149	123.5	92.5	84.5	89.0	86
YOLOv3-tiny	866	17.4	88.1	82.8	87.5	714
YOLOv5l	4610	92.8	87.0	82.8	88.1	119
YOLOv5m	2085	42.2	89.0	81.5	89.0	145
YOLOv5x	8617	173.1	84.5	82.8	88.6	79
YOLOv5s	701	14.4	82.5	84.7	87.1	256
YOLOv5n	176	3.9	84.0	81.2	86.0	212
YOLOv5s-MESB(本文方法)	742	15.3	87.7	83.3	89.3	227

从实验结果可以看出，相较于其他几种目标检测算法，

本文提出的算法检测精度最高。其中, YOLOv3、YOLOv5m以及YOLOv5x的检测精度与本文提出的算法检测精度较为相近,但本文所提算法权重文件大小只有YOLOv3的12.3%,只有YOLOv5m的36.2%,只有YOLOv5x的8.8%,体量优势明显;相较于检测速度较为相近的YOLOv5n算法以及YOLOv5l算法,本文所提算法的检测精度优势明显,比YOLOv5n高3.3%,比YOLOv5l高1.2%,此外,虽然YOLOv5n有最小的权重文件,仅有3.9 MB,但是检测精度却较低,只有86%,无法精确的识别有无人员吸烟,相较于权重文件大小相近的YOLOv5s、YOLOv3-tiny,本文提出算法的检测精度分别提高了2.2%、1.8%。综上所述,本文提出的YOLOv5s-MESB算法在保持轻量化的同时,有着最高的检测精度,并且保持着较好的实时性,整体表现较为突出,证明了本文所提算法的优越性。

#### 4 结束语

由于距离远,从视频监控中捕捉到的吸烟人员相对较小,图像相对模糊,用原始的YOLOv5s网络检测吸烟人员精度较差,并且误检漏检严重,为此本文通过对多头自注意力层、ECA注意力模块、损失函数Siou、BiFPN等方法的研究,将其统一改进到YOLOv5s网络中,进一步优化YOLOv5s算法。实验结果证明,改进后的算法对于吸烟行为的检测精度提高了2.2%,检测速度达到了227FPS,在保持算法速度和轻量化的同时获得了更高的检测精度,相比其他目标检测算法,更加适用于对吸烟行为的检测。但是,对于人员密集或是在阴暗的场合,本文改进的算法会存在一些误检,影响检测效果,因此本文算法还有很大的改进空间,后续工作可以针对提高密集人群和阴暗场景下吸烟检测精度进行研究,来进一步提升吸烟行为的检测效果。

#### 参考文献:

- [1] 杨若聪. 基于YOLO算法的火电厂人员行为安全识别 [D]. 北京: 华北电力大学, 2021.
- [2] 李芬. 智能烟火报警器的设计与实现 [J]. 电脑知识与技术, 2020, 16 (2): 238-239.
- [3] 刘阳, 谢永强, 李忠博, 等. 基于深度学习的目标检测算法研究进展 [J]. 通信技术, 2021, 54 (9): 2063-2073.
- [4] 贺宇哲, 何宁, 张人, 等. 基于双阶段目标检测算法研究综述 [C] //中国计算机用户协会网络应用分会2021年第二十五届网络新技术与应用年会论文集, 2021: 182-186.
- [5] LIU W, ANGUELOV D, ERHAN D. SSD: Single shot multibox detector [C] //European conference on computer vision. Springer, Cham, 2016: 21-37.
- [6] 宋汝鹏. 单阶段目标检测算法研究与运用 [D]. 成都: 电子科技大学, 2021.
- [7] 张瑶, 卢焕章, 张路平, 等. 基于深度学习的视觉多目标跟踪算法综述 [J]. 计算机工程与应用, 2021, 57 (13): 55-66.
- [8] 韩桂金, 李倩. 基于Faster R-CNN的吸烟快速检测算法 [J]. 西安邮电大学学报, 2020, 25 (2): 85-91.
- [9] 王梦依. 基于深度学习图像处理的吸烟行为检测 [D]. 北京: 北方工业大学, 2020.
- [10] 姜晓凤, 王保栋, 夏英杰, 等. 基于人体关键点和YOLOv4的吸烟行为检测 [J]. 陕西师范大学学报 (自然科学版), 2022, 50 (3): 96-103.
- [11] 杨雨, 陈服, 陈柏青, 等. 基于深度学习的变电站保护硬压板检测与状态识别技术研究 [J]. 科技创新与应用, 2021, 11 (24): 25-29.
- [12] 陈耀祖, 谷玉海, 成霞, 等. 基于优化YOLOv4算法的行驶车辆要素检测方法 [J]. 应用光学, 2022, 43 (2): 248-256.
- [13] SRINIVAS A, LIONT Y, PARMAR N. Bottleneck Transformers for Visual Recognition [J]. arXiv e-prints, 2021, arXiv: 2101.11605.
- [14] 胡哲, 徐承志, 雷光波, 等. 基于改进YOLOv5的X射线图像铸件缺陷实时检测 [J]. 激光杂志, 2022, 43 (5): 54-59.
- [15] WANG Q, WU B, ZHU P. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks [C] //2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020.
- [16] 华泽玺, 施会斌, 罗彦, 等. 基于轻量级YOLO-v4模型的变电站数字仪表检测识别 [J/OL]. 西南交通大学学报: 1-11 [2022-07-22]. <https://cc0eb1c56d2d940cf2d0186445b0c858cnki.casb4.njit.edu.cn:5443/kcms/detail/51.1277.U.20211027.1050.003.html>.
- [17] 周旗开, 张伟, 李东锦, 等. 基于改进YOLOv5s的光学遥感图像舰船分类检测方法 [J/OL]. 激光与光电子学进展: 1-14 [2022-07-22]. <https://cc0eb1c56d2d940cf2d0186445b0c858cnki.casb4.njit.edu.cn:5443/kcms/detail/31.1690.TN.20220714.1336.437.html>.
- [18] 李功, 赵巍, 刘鹏, 等. 一种用于目标跟踪边界框回归的光滑IoU损失 [J]. 自动化学报, 2023, 49 (2): 288-306.
- [19] ZHORA GEVORGYAN. Siou Loss: More Powerful Learning for Bounding Box Regression [J]. arXiv e-prints, 2022, arXiv: 2205.12740.
- [20] TAN M, PANG R, LE Q V. EfficientDet: Scalable and Efficient Object Detection [C] //2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.
- [21] 杨文涛, 张维光. 基于改进YOLOv5m的弱小目标识别方法 [J/OL]. 计算机测量与控制: 1-9 [2022-09-20]. <http://kns.cnki.net/kcms/detail/11.4762.TP.20220818.1008.014.html>.
- [22] 王鹏飞, 黄汉明, 王梦琪. 改进YOLOv5的复杂道路目标检测算法 [J/OL]. 计算机工程与应用: 1-13 [2022-07-22]. <https://cc0eb1c56d2d940cf2d0186445b0c858cnki.casb4.njit.edu.cn:5443/kcms/detail/11.2127.TP.20220623.1640.012.html>.